

This listing of claims will replace all prior versions, and listings, of claims in the application:

Claims 1-45 (deleted)

1 Claim 46 (currently amended): A method for filtering  
2 search results to remove near-duplicates, the method  
3 comprising:  
4 a) for each of a plurality ~~predetermined number~~ of  
5 candidate search results, determining whether the  
6 candidate search result is a near-duplicate of another  
7 candidate search result by  
8 1) comparing a cluster identifier of the  
9 candidate search result with that of the other  
10 candidate search result, and  
11 2) if the cluster identifiers of the two  
12 candidate search results match, then concluding  
13 that the two candidate search results are  
14 near-duplicates; and  
15 b) if it is determined that the candidate search  
16 result is a near-duplicate of the other ~~another~~  
17 candidate search result, then rejecting the candidate  
18 search result.

1 Claim 47 (currently amended): A search filter for  
2 processing search results to remove near-duplicates, the  
3 search filter comprising:  
4 a) a near-duplicate determination facility for  
5 determining, for each of a plurality ~~predetermined~~  
6 ~~number~~ of candidate search results, whether the  
7 candidate search result is a near-duplicate of another  
8 candidate search result, and wherein the

9 near-duplicate determination facility includes a  
10 comparison facility for comparing a cluster identifier  
11 of the candidate search result with that of another  
12 candidate search result, and wherein if the cluster  
13 identifiers of the two candidate search results match,  
14 then it is concluded that the two candidate search  
15 results are near-duplicates; and  
16 b) a filter for rejecting the candidate search result  
17 if it is determined that the candidate search result  
18 is a near-duplicate of the other ~~another~~ candidate  
19 search result.

1 Claim 48 (currently amended): A machine-readable medium  
2 having stored thereon a plurality of records, each of the  
3 records comprising:  
4 a) a first field for storing a document identifier;  
5 and  
6 b) a plurality of lists, each of the plurality of  
7 lists containing elements of a document identified by  
8 the document identifier stored in the first field,  
9 wherein a hash function is used to hash each of  
10 the elements in order to determine which one of the  
11 plurality of lists that each of the elements will be  
12 contained in.

1 Claim 49 (currently amended): A method for determining  
2 whether two documents are near-duplicates, the method  
3 comprising:  
4 a) for each of the two documents, generating at least  
5 two fingerprints; and  
6 b) determining whether or not the two documents are  
7 near-duplicate documents by

8           1) determining whether or not any one of the at  
9           least two fingerprints of a first of the two  
10          documents matches any one of the at least two  
11          fingerprints of a second of the two documents,  
12          and  
13          2) if it is determined that any one fingerprint  
14          of the at least two fingerprints of the first of  
15          the two documents does match any one fingerprint  
16          of the at least two fingerprints of the second of  
17          the two documents, then concluding that the two  
18          documents are near-duplicates.

1   Claim 50 (previously presented): A machine-readable  
2   medium having stored thereon a plurality of records, each  
3   of the records comprising:

- 4       a) a first field for storing a document identifier;  
5       and  
6       b) a plurality of lists, each of the plurality of  
7       lists containing elements of a document identified by  
8       the document identifier stored in the first field,  
9       wherein at least some of the plurality of lists  
10      include different numbers of elements.

1   Claim 51 (previously presented) The machine-readable  
2   medium of claim 51 wherein at least one of the plurality of  
3   lists include no elements.

1   Claim 52 (previously presented): A machine-readable  
2   medium having stored thereon a plurality of records, each  
3   of the records comprising:

- 4   a) a first field for storing a document identifier; and

5 b) a plurality of lists, each of the plurality of lists  
6 containing elements of a document identified by the  
7 document identifier stored in the first field,  
8 wherein contiguous elements in a document are not  
9 necessarily contiguous elements of a list.

1 Claim 53 (previously presented): A machine-readable  
2 medium having stored thereon a plurality of records, each  
3 of the records comprising:

4 a) a first field for storing a document identifier; and  
5 b) a plurality of lists, each of the plurality of lists  
6 containing elements of a document identified by the  
7 document identifier stored in the first field,  
8 wherein for each of the records, the number of  
9 lists is the same.

1 Claim 54 (currently amended): The machine-readable medium  
2 of claim 53 ~~54~~ wherein a number of the plurality of lists  
3 is independent of document size.

1 Claim 55 (new): The machine-readable medium of claim 48  
2 wherein each of the elements of a document is an element  
3 that has been extracted from the document.

1 Claim 56 (new): The machine-readable medium of claim 48  
2 wherein each of the elements of a document is a  
3 predetermined one of (A) a predetermined number of words,  
4 (B) a predetermined number of sentences, (C) a  
5 predetermined number of characters, (D) a predetermined  
6 number of paragraphs, and (E) a predetermined number of  
7 sections.

1 Claim 57 (new): The machine-readable medium of claim 48  
2 wherein each of the elements of a document partially  
3 overlaps another of the elements of the document.

1 Claim 58 (new): The machine-readable medium of claim 50  
2 wherein each of the elements of a document is an element  
3 that has been extracted from the document.

1 Claim 59 (new): The machine-readable medium of claim 50  
2 wherein each of the elements of a document is a  
3 predetermined one of (A) a predetermined number of words,  
4 (B) a predetermined number of sentences, (C) a  
5 predetermined number of characters, (D) a predetermined  
6 number of paragraphs, and (E) a predetermined number of  
7 sections.

1 Claim 60 (new): The machine-readable medium of claim 50  
2 wherein each of the elements of a document partially  
3 overlaps another of the elements of the document.

1 Claim 61 (new): The machine-readable medium of claim 52  
2 wherein each of the elements of a document is an element  
3 that has been extracted from the document.

1 Claim 62 (new): The machine-readable medium of claim 52  
2 wherein each of the elements of a document is a  
3 predetermined one of (A) a predetermined number of words,  
4 (B) a predetermined number of sentences, (C) a  
5 predetermined number of characters, (D) a predetermined  
6 number of paragraphs, and (E) a predetermined number of  
7 sections.

1 Claim 63 (new): The machine-readable medium of claim 52  
2 wherein each of the elements of a document partially  
3 overlaps another of the elements of the document.

1 Claim 64 (new): The machine-readable medium of claim 53  
2 wherein each of the elements of a document is an element  
3 that has been extracted from the document.

1 Claim 65 (new): The machine-readable medium of claim 53  
2 wherein each of the elements of a document is a  
3 predetermined one of (A) a predetermined number of words,  
4 (B) a predetermined number of sentences, (C) a  
5 predetermined number of characters, (D) a predetermined  
6 number of paragraphs, and (E) a predetermined number of  
7 sections.

1 Claim 66 (new): The machine-readable medium of claim 53  
2 wherein each of the elements of a document partially  
3 overlaps another of the elements of the document.

1 Claim 67 (new): A machine-readable medium having stored  
2 thereon a plurality of records, each of the records  
3 comprising:  
4 a) a first field for storing a document identifier;  
5 and  
6 b) a plurality of lists, each of the plurality of  
7 lists containing elements of a document identified by  
8 the document identifier stored in the first field,  
9 wherein each of the elements are contained in one of  
10 the plurality of lists in accordance with a result of  
11 hashing the element using a hash function.